Cluster 1 during this period (1987–1991) describes research topics relating to online catalogues, data storage, office automation, optical data storage and multiprocessor systems, and so on. Cluster 2 describes research topics relating to natural language processing, linguistic analysis, neural models, library and information science, memory, and so on. Cluster 3 describes topics on user services, cataloguing, storage, bibliographic databases, programming, and so on. Cluster 4 focuses on information storage and retrieval, databases, science and technology, parallel processing, information services, and so on. Finally, Cluster 5 appears to focus on information work, subject indexing, computerized information storage and retrieval, technical services, searching, and so on.

As shown in Table 2, during this period, for each cluster, the outer links among keywords are much more than its inner links. Around 88% of keywords in each cluster have outer links with other keywords based on the Salton Index. These indicate that the links among keywords based on the Salton Index do not aggregate within the cluster. In other words, around 62% of such kind of links are located among different clusters. So, these links not only reflect the abundant relationships among clusters, but also show a loosely internal composition in each cluster. But, the average link per keyword is very high during this period. It strongly suggests
the IR community and it is able to maintain itself to develop over the course of time in IR field.


During this period, among 240 keywords, only one keyword (i.e., neural models) did not appear so that this keyword was excluded from this period of research. Thus, 239 keywords were chosen as the keyword research sample in this period. The same method was used to generate the general overview map of the IR field in 1992–1997 by MDS (Fig. 13) and each cluster (sub-domain) was labeled by the most frequent keywords within the cluster as before.

Each cluster contains around 50 keywords. In order to construct ‘fine-structure’ or detailed maps as well, each cluster was chosen as the input variable to map the sub-domain based on $239 \times 239$ correlation matrix. Five detailed sub-domain maps (Figs. 14–18) were generated to reflect specific characters of each sub-domain in IR field.

Cluster 1 describes research topics relating to information storage and retrieval, searching, systems analysis, online information retrieval, database, and so on. Cluster 2 during this period describes research topics relating to networks, multimedia, WWW, medicine, Internet, and so on. Cluster 3 mainly describes topics on data storage, optical data storage, magnetic data storage,

Ying Ding, Gobinda Chowdhury, Schubert Foo
Division of Information Studies, School of Applied Science
Nanyang Technological University, Nanyang Avenue,
Singapore 639798

Abstract

Author cocitation analysis (ACA) has been used to exploit changes in the intellectual base of information retrieval (IR) field over two consecutive time periods, 1987-1991 and 1992-1997. Thirty nine highly cited IR researchers are selected as the research sample. Multidimensional scaling (MDS) and clustering techniques (CT) were used to create the two-dimensional maps to display the dynamic intellectual structure of IR based on scholars citing their work over these two time periods. Factor analysis (FA) has been used to reveal the “breath” of the authors’ research areas.

ACA offers a good technique that contributes to the understanding of intellectual structure in the sciences and possibly in other areas to the extent that those areas rely on formal scholarly communication such as serial publications. Nonetheless, obvious drawbacks exist in ACA. These include the subjective nature of the interpretation of results, the difficulty of readily identifying clusters, and the inability to distinguish collaborative research relationships between authors. Thus, ACA by itself is insufficient. However, ACA can be enhanced significantly when combined with FA to give a more accurate and useful picture of the MDS results.

Keywords: Author cocitation analysis, information retrieval, multidimensional scaling, factor analysis, intellectual structure.

Introduction

As the structure of our society becomes increasingly complex, individuals and groups need to communicate an increasing volume of information. Schramm [1], a dominant figure in the communication sciences, once said that the field of communication was like “an oasis in the desert, where many trails cross, and many travellers pass, but only a few tarry”. Bibliometrics or “the application of mathematics and statistical methods to books and other media of communication” provide a method for examining communication among scholars in a field through their scholarly publication [2].
underlying their placement. Figures 1 and 2 show the two-dimensional MDS mapping generated by the ALSCAL program (part of SPSS [21]) with good fit for the two time periods.

Clustering techniques (CT) are used to group authors so as to provide insights into the intellectual organisation of a given field. The cluster generating programs are also available in SPSS. Authors are grouped within boundaries, as shown in solid lines. The name of the cluster was chosen based on the common research topics of each author in this cluster (The details of MDS and CT are discussed in detail in references [21, 22] and [23] and will not be repeated here).

**General structural stability**

In general, there are some schools of thought which both appear in the two periods’ ACA map, such as general IR theory, IR model, IR techniques, information seeking and retrieving behaviour and so on.

![Figure 1. Author cocitation mapping of IR (1987-1991)](image-url)
General IR theory, whose main sole contributor is Salton, is distinct and located on the right hand of both the maps. This result coincides with that of another research [24]. In contrast, user information seeking and retrieving behavioural research of Figure 1 has been divided more distinctly into user perspectives of IR, online information seeking & retrieving behavioural, and information seeking and retrieving model (user searching strategies) in Figure 2. All these groups are located together or in the middle of the map. This reflects that they are vital research areas in IR research.

The horizontal axis (from left to right) of 11-year period seems to represent more specific and technological researches to more general and basic, theoretical researches. The vertical axis (from top to bottom) for the same period appears to see a shift from theoretical system design to application and evaluation, and finally to user searching behaviour.

Scholarly migration

The general spatial orientation of authors and their cluster assignment, have not changed much in these two periods. However, no single author has maintained exactly the same position in the map during these two time periods. This does not imply that all authors are moving to new research areas. Even if an author’s research area remains, the position of that author in different maps might change since these maps were derived on the citation relationships of all the authors. For example, Sparck Jones’s position in these two maps is very different, one is above the horizontal axis and one is below the horizontal axis although her research areas did not change much.

On the other hand, when a comparison of the position of an author is made in these two maps, we can see that although some authors’ positions did not change, the research groups